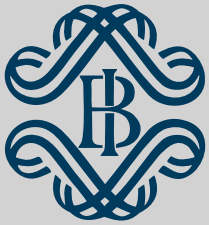
BANCA D'ITALIA

Temi di discussione

del Servizio Studi

Non nested testing procedures: Monte Carlo evidence and post simulation analysis in dynamic models

by Giuseppe Parigi



Numero 106 - Ottobre 1988

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

Non nested testing procedures: Monte Carlo evidence and post simulation analysis in dynamic models

by Giuseppe Parigi

Numero 106 - Ottobre 1988

The purpose of the «Temi di discussione» series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

Abstract

The purpose of this paper is to describe some of the widely used non nested testing procedure such as the Cox-type family tests, the F-test and two new procedures recently proposed by Godfrey-Pesaran.

The general framework used for the analysis of these procedures is the encompassing principle which allows to classify all them into two groups: variance-encompassing and parameter-encompassing tests.

As the small sample behaviour of most of these tests is not known, Monte Carlo experimentation is required to assess the validity of different procedures. Power and size comparisons are then performed in the context of alternative dynamic models characterized by equal and unequal number of regressors. Monte Carlo simulations are however limited in the sense that they do not cover the number of all possible experiments. In this way, the results of such techniques are imprecise and characterized by specificity. To avoid this problem a possible solution is to resort to the application of response surface techniques. After a brief exposition of the main features of response surface analyses, an application is shown for the F-test and one of the Cox-type testing procedures derived by Godfrey-Pesaran.

1. INTRODUCTION (*)

It is a common fact that a great deal of economic phenomena can be interpreted according to different and often conflicting theories. A crucial role is then assigned to econometric techniques in order to find evidence in favour of one particular specification. As the true Data Generating Process (DGP) is not known - nor it is possible to know - the only output of econometric research is then a set of approximations which may not be directly comparable from the point of view of classical testing procedures. This is precisely the case of non nested models, or models that can not be obtained one from the other by imposing a set of parametric restrictions. In presence of more alternative non nested formulations two different problems have to be distinguished: the problem of model discrimination and the problem of model specification. In the first case, one model is always chosen as the best among different alternatives, while in the second case the specification of a model is checked by using the performance of competing models. In both cases, being the models non nested it is not possible to employ any of the classical testing procedures based on the likelihood principle as their distribution is unknown. Clearly, a direct solution could be to derive analytically or numerically the distribution of particular testing procedures in each case, but this is inefficient and very expensive.

(*) This paper was written in September 1986 at the University of Warwick where the author was attending the M.A. course in economics. He would like to thank K.F. Wallis, W. Narendranathan and M. Salmon for their many useful comments and helpful advice. However, the usual disclaimer applies.

To solve this problem two different approaches are currently used in the econometric practice. The first one, directly concerned with the problem of discriminating among separate models, relies on the optimization of criteria based on the standard error of the regression (adjusted R^2, Akaike Information Criterium, etc.). According to this approach a model is always chosen as the preferred one by considering its fit along with the parsimony of its specification. Apart from the "ad hoc" nature of these criteria (there does not exist a statistical theory underlying the comparison of the R^2 of two or more non nested models) the main criticism concerns the fact that the problem of discrimination is solved by considering only the performance of a model without taking into account its capacity to predict the performance of alternative specifications.

The last observation forms the basis of a second and more fruitful approach: the encompassing principle. In this context, if a model, say M_0, is assumed to be the best approximation to the DGP (the null hypothesis) it has to predict the performance of the alternative(s). The testing procedure is then obtained by comparing the actual performance of the alternative with the performance that could be expected were M_0 supposed to be mimicking the DGP. This is the encompassing approach as described by Mizon (1984) and Mizon-Richard (1986). Basically, encompassing tests may be considered as model specification tests because a rejection of the null hypothesis does not provide evidence in favour of the alternative. In this context, the alternative model(s) is (are) used as a guide in testing the specification of the "true" model. However such tests may also be used in a discriminating context when they are employed in a particular testing procedure defined as "paired separate tests" (see McAleer (1984)). In this case, given two competing models, M_0 and M_1, two test-statistics are computed: one with M_0 as the null hypothesis and M_1 as the alternative; the other with the role of the two models reversed. There are four possible outcomes deriving from the application of paired separate tests: 1) accept M_0; 2) accept M_1; 3) accept both models; 4) reject both models. Discrimination is then possible only in the

case 1) and 2) (it is worth pointing out that the application of paired separate tests is justified if the two models, M_0 and M_1 are meaningful). In cases 3) and 4), the outcome is clearly indeterminate so the only possible suggestion is that the two models are somehow misspecified.

In practice there are different kinds of encompassing test-statistics according to the statistic used to measure the performance of a model. Following the terminology used by Mizon (1984) and Mizon-Richard (1986) the whole set of non nested encompassing tests may be classified into two groups: variance encompassing and parameter encompassing tests. The former relies on the variance of the disturbance term and includes the family of Cox-type tests - i.e. Cox tests, J and JA tests - while the latter, based on the whole set of parameters, concerns the ortodox or comprehensive model approach characterized by the use of the F test. Sections 2 and 3 review some of the most important features of these tests; in particular, section 3 contains an analysis of the asymptotic properties, consistency, efficiency, etc. of the different testing procedures. As all these properties are equivalent for all tests a possible discrimination between them may be based on the small sample behaviour of the tests. In this context, problems arise in connection with the Cox-types tests and the J test because their exact distribution is not known. Moreover, when some form of dynamics is introduced into the analysis, then, even those procedures which are exact in the static case - i.e. the F test and the JA test - are valid only asymptotically. The only possibility of deriving some information about small sample behaviour is therefore the application of Monte Carlo techniques. Some of the most important results obtained by Monte Carlo simulation is briefly summarized in section 4. This section also contains a description of two different small sample adjusted Cox tests derived by Godfrey-Pesaran (1983) (GP hereafter).

Monte Carlo experimentation is not free of problems. Like all statistical experimentation it is characterized by some drawbacks that should be taken into account by the researcher. The two main problems connected with the implementation of Monte Carlo

analysis are the imprecision and the specificity of the results (see Cochran-Cox (1957), Mizon-Hendry (1980) and Hendry (1984)). In order to limit the possible effects of these two negative characteristics most researchers have produced a great number of experiments with the consequence of publishing an increasing quantity of tables and numbers. As it can be seen, this kind of solution lacks efficiency from the point of view both of the analysis and of the potential reader. In this context, the response surface analysis technique may provide a viable alternative as Mizon-Hendry (1980) and Ericsson (1986) show. A response surface analysis or, in Ericsson's terminology, post simulation analysis, is applied to a particular case of GP's study of the small sample behaviour of non nested testing procedures. In particular, we consider the case where both models, the data generation process and the alternative, are dynamic in the sense that they contain a lagged dependent variable. Moreover, the case where the two models have a different number of regressors is investigated as well. Section 5 describes the experimental design, while section 6 illustrates the methodology employed in the simulation exercise along with a brief analysis of the main characteristics of response surface techniques.

Section 7 illustrates the results of the Monte Carlo simulation and the consequent application of response surface analysis to the F test and one of the small sample adjusted Cox test proposed by GP. The estimates of power and size we obtain are different from what GP present. This is essentially due to the fact that our analysis is characterized by the presence of an intercept term in both competing models, whereas GP do not take into account any constants. The results of response surface analysis yield an indication of the potentialities of such an approach to evaluate Monte Carlo simulations.

Some concluding remarks and suggestions about further research in this area are finally reported in section 8.

2. TESTING PROCEDURES

The encompassing principle relies on the comparison of the actual performance of an alternative model with its expected performance under the null hypothesis given by the model supposed to be the best approximation to the unknown DGP. In other words, the two different performances of the alternative model must not be significantly different if the null hypothesis is true. In order to give a formal expression to this concept throughout all our analysis we concentrate exclusively upon the linear regression model, although it is possible to extend the results to the non linear case. The choice of the linear case is essentially motivated by the simplicity of the calculations involved.

The general set up characterizing our investigation is given by the two non nested linear regression models:

$$[2.1] \qquad M_0 : y = Xb_0 + U_0$$

$$[2.2] \qquad M_1 : y = Zb_1 + U_1$$

where y is the (Tx1) vector of observations on the dependent variable; X and Z are the (Txk_0) and (Txk_1) non stochastic observation matrices; b_0 and b_1 are the (k_0x1) and (k_1x1) parameter vectors and U_0 and U_1 are the (Tx1) vector of disturbances. It is assumed that the U's are distributed as $NID(0,\sigma_i^2 I_T)$ where i = 0,1. It is customary to assume that the matrices (X'X/T) and (Z'Z/T) converge to well defined positive limits and that the limit of Z'X/T does not vanish. M_0 and M_1 are supposed to be non nested in the sense that the columns of Z (X) can not be written as linear combination of the columns of X (Z). In this context, the structure of a test for separate models need assume a model as "true" (the null hypothesis or M_0), and the other as the alternative (M_1). According to Mizon (1984) and Mizon-Richard (1986) M_0 is said to encompass M_1 if and only if:

$$[2.3] \qquad D = d_1 - d_{10} = 0$$

where d_1 is a statistic relevant to the analysis of M_1 - in our case, d_1 is the whole parameter set of M_1, $d_1 = (b_1, \sigma_1^2)$; d_{10} is the expectation of d_1 under the null hypothesis given by M_0. For practical purposes, d_1 and d_{10} may be replaced by their maximum likelihood estimators, $\hat{d}_1$ and $\hat{d}_{10}$.

The first step in the construction of an efficient testing procedure is the derivation of the limiting distribution of $\sqrt{T}(\hat{D})$, which Mizon ((1984), page 149) shows to be:

$$[2.4] \qquad \sqrt{T}(\hat{D}) = \sqrt{T}(\hat{d}_1 - \hat{d}_{10}) \xrightarrow[M_0]{d} N(0, V_0(\hat{d}_1 - \hat{d}_{10}))$$

where $V_0(\hat{d}_1 - \hat{d}_{10}) = V_0(\hat{d}_1) - V(\hat{d}_{10})$.

This result is of great importance because it allows to obtain a Wald test for separate models. In effect, according to the statistic of interest we choose, it is possible to derive different Wald tests. For instance, if $\hat{\sigma}_1^2$ is chosen, then we have (for the complete derivation of n_1 and n_2, see Mizon (1984) and/or Mizon-Richard (1986)):

$$[2.5] \qquad n_1 = \frac{T(\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2)^2}{V_0(\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2)} \sim \chi^2(1)$$

In the same way, if b_1 is chosen:

$$[2.6] \qquad n_2 = T(\hat{b}_1 - \hat{b}_{10})' V_0(\hat{b}_1 - \hat{b}_{10})^{-1} (\hat{b}_1 - \hat{b}_{10}) \sim \chi^2(r)$$

where r is the number of non overlapping variables in the matrix Z. As n_1 is based on the error variance of M_1 it is defined as a variance encompassing test while n_2 as a parameter encompassing test. Mizon-Richard show that as all the Cox-type testing procedures are variance encompassing tests they are asymptotic equivalent to n_1. This is the reason why these tests are called one degree of freedom tests. On the other hand, n_2 is shown to be

asymptotic equivalent to the classical F-test for separate models applied in the context of the orthodox or comprehensive model approach.

In this section, we first analyze some variance encompassing tests such as the Cox test and a linear version of it; the J and the JA tests. Secondly, for the parameter encompassing tests the widely used F-test is considered.

2.1 Variance encompassing tests

2.1.1 The Cox test

The Cox test is based on a modified likelihood ratio test (LR hereafter) as suggested by Cox (1961), (1962). Given the LR for the models [2.1] and [2.2]:

$$LR = L_0 - L_1$$

where L_0 and L_1 are the maximized log-likelihood functions of M_1 (i=0,1), if M_0 were nested in M_1, then the asymptotic expectation of LR, evaluated under the null, would be zero. This does not occur with non nested models so Cox proposed to correct LR by subtracting its asymptotic mean. Thus the Cox text is based upon:

$$[2.7] \qquad T_0 = (L_0-L_1) - T \cdot [\text{plim}_0 T^{-1}(L_0-L_1)]_{\mu=\hat{\mu}_0}$$

where plim_0 denotes probability limit under M_0 and $\mu_i = (b_i, \sigma^2_i)$ (i=0,1) is the vector of parameters characterizing the two models. Cox proves under general conditions that if V_0 is the variance of T_0, then (see Cox (1961)):

$$[2.8] \qquad N_0 = T_0/(V_0)^{\frac{1}{2}} \sim N(0,1)$$

where ~ means "converges in distribution to" (this result is not affected if we use a consistent estimator, say $\hat{V}_0$, instead of the

true variance).

In the linear case, a formal expression for T_0 may be derived by observing that:

$$[2.9] \qquad L_1 = -(T/2)[1+\log(2\pi\hat{\sigma}_i^2)] \quad (i = 1,0)$$

where $\hat{\sigma}_i^2$ (i=0,1) is the maximum likelihood estimate of σ_i^2; and

$$T[\text{plim}_0 T^{-1}(L_0)]_{\mu=\hat{\mu}_0} = L_0$$

T_0 may then be rewritten as:

$$[2.10] \qquad T_0 = -L_1+T[\text{plim}_0 T^{-1}(L_1)]_{\mu=\hat{\mu}_0} = T/2\log(\hat{\sigma}_1^2/\hat{\sigma}_{10}^2)$$

where $\hat{\sigma}_{10}^2$ denotes the plim of $\hat{\sigma}_1^2$ under H_0, that is for $\mu=\hat{\mu}_0$. To derive $\hat{\sigma}_{10}^2$ we may notice that:

$$[2.11] \qquad \hat{\sigma}_1^2 = 1/T(y-Z\hat{b}_1)'(y-Z\hat{b}_1) =$$

$$= 1/T(y-X\hat{b}_0+X\hat{b}_0-Z\hat{b}_1)'(y-X\hat{b}_0+X\hat{b}_0-Z\hat{b}_1)$$

$$= 1/T(y-X\hat{b}_0)'(y-X\hat{b}_0)+2(y-X\hat{b}_0)'(X\hat{b}_0-Z\hat{b}_1)+(X\hat{b}_0-Z\hat{b}_1)'(X\hat{b}_0-Z\hat{b}_1)]$$

where $\hat{b}_0$ and $\hat{b}_1$ are $(X'X)^{-1}X'y$ and $(Z'Z)^{-1}Z'y$ respectively.

Under the null hypothesis that M_0 is the true specification, $(y-X\hat{b}_0)=\hat{e}_0$ is asymptotically uncorrelated with $(X\hat{b}_0-Z\hat{b}_1)$, that is

$$\text{plim}_0\ \hat{e}_0'(X\hat{b}_0-Z\hat{b}_1) = \emptyset$$

This comes from the fact that $\hat{e}_0$ tends to a stochastic variable with mean zero and variance σ_0^2, while $(X\hat{b}_0-Z\hat{b}_1)$ to a non stochastic limit (see Davidson-McKinnon (1981)). In particular,

$$\text{plim}_0 Z\hat{b}_1 = Z(Z'Z)^{-1}Z'Xb_0$$

so from [2.11]:

$$[2.12] \quad \text{plim}_0 \hat{\sigma}_1^2 = \sigma_{10}^2 = \sigma_0^2 + \lim(1/T)[b_0'X'(I-Z(Z'Z)^{-1}Z')Xb_0]$$

by replacing the elements of σ_{10}^2 with their consistent estimates, under $\mu=\hat{\mu}_0$, we get:

$$\hat{\sigma}_{10}^2 = \hat{\sigma}_0^2 + (1/T)[\hat{b}_0'X'(I-Z(Z'Z)^{-1}Z')X\hat{b}_0]$$

Thus $\hat{\sigma}_{10}^2$ may be seen as the sum of two components: an estimate of the variance of the "true" model and an estimate of the additional variance due to the difference between M_0 and M_1.

As a last step in the derivation of N_0, we need an estimate of V_0. This may be obtained by applying a result of Cox (1962) (see also Pesaran (1974)) such that:

$$[2.13] \quad \hat{V}_0 = (\hat{\sigma}_0^2/\hat{\sigma}_{10}^4)\ (\hat{b}_0'X'A_1A_0A_1X\hat{b}_0)$$

where $A_0=(I-X(X'X)^{-1}X')$ and $A_1=(I-Z(Z'Z)^{-1}Z')$.

N_0 is finally given by:

$$[2.14] \quad N_0 = T/2\log(\hat{\sigma}_1^2/\hat{\sigma}_{10}^2)/[(\hat{\sigma}_0^2/\hat{\sigma}_{10}^4)(\hat{b}_0'X'A_1A_0A_1X\hat{b}_0)]^{\frac{1}{2}}$$

It may seem extremely cumbersome to compute N_0, but it can be easily obtained by the means of four regressions. In effect, we have:

$$[2.15] \quad N_0 = \frac{T}{2}\log\frac{e_1'e_1}{e_0'e_0 + e_{10}'e_{10}} \Big/ \frac{[(e_0'e_0)(e_{100}'e_{100})]^{\frac{1}{2}}}{(e_0'e_0 + e_{10}'e_{10})}$$

where e_i ($i=0,1$) is the OLS residual vector of the regression of M_0 and M_1;

e_{10} is the OLS residual vector of the regression of

$\hat{y}_0 (=X\hat{b}_0)$ on Z;

e_{100} is the OLS residual vector of the regression of e_{10} on X

(see Pesaran (1974); for the non linear case and systems of equations see Pesaran and Deaton (1978)).

By noting that

$$T/2\log(\hat{\sigma}_1^2/\hat{\sigma}_{10}^2) \simeq T/2\log(\hat{\sigma}_1^2/\hat{\sigma}_{10}^2 - 1)$$

Fisher-McAleer (1981) show that [2.15] is asymptotically equivalent to

$$[2.16] \qquad NL_0 = T/2\log(\hat{\sigma}_1^2-\hat{\sigma}_{10}^2)/\{(\hat{\sigma}_0^2/\hat{\sigma}_{10}^4)(\hat{b}_0'X'A_1A_0A_1X\hat{b}_0)\}^{\frac{1}{2}}$$

[2.16] is useful because it shows that the value of the Cox statistic crucially depends on the sign of $(\hat{\sigma}_1^2-\hat{\sigma}_{10}^2)$. Specifically,

$$(\hat{\sigma}_1^2-\hat{\sigma}_{10}^2) \gtreqless 0 \qquad \text{as} \qquad NL_0 \gtreqless 0$$

so that a significant negative Cox-test means that the actual performance of M_1 is better than expected and therefore H_0 must be rejected. Similarly a significant positive value of NL_0 implies that the alternative is performing worse than expected and H_0 must be rejected as well. H_0 is not rejected only when

$$\sigma_1^2 - \sigma_{10}^2 \simeq 0$$

In this case the performance of M_1 under H_0 is as expected and we may reasonably suppose that M_0 variance-encompasses M_1.

2.1.2 The J and the JA test

One of the main drawbacks of the Cox test, as shown in [2.15] or [2.16], is that it is very cumbersome to derive. In order to simplify the computation, several different testing procedures have been proposed in the literature. Two of these test

statistics are the J test of Davidson-McKinnon (1981) and the JA test of Fisher-McAleer (1981) which can be easily obtained by the means of an auxiliary regression. In the preceding paragraph it has been shown that the Cox procedure is based upon the comparison of $\hat{\sigma}_1^2$ and $\hat{\sigma}_{10}^2$. Specifically, the Cox test checks whether:

$$[2.18] \qquad (y-X\hat{b}_0)'(X\hat{b}_0-Z\hat{b}_1) = 0$$

In other words, it is checked whether the residuals from the "true" model are uncorrelated with the difference between the fitted values from M_0 and M_1.

In this context, a simpler way to test [2.18] is by the means of the test of α in the regression:

$$[2.19] \qquad y = (1-\alpha)X\hat{b}_0 + \alpha Z\hat{b}_1 + w$$

As such test -defined as the C test - is conditional on $\hat{b}_0$ and $\hat{b}_1$ the derivation of its distribution is quite complex, so Davidson and McKinnon (1981) suggest to resort to the comprehensive model:

$$[2.20] \qquad y = (1-\alpha)Xb_0 + \alpha Zb_1 + w$$

In [2.20] α, b_0 and b_1 are not identified, so this model can not be used. However, as $\hat{b}_1$ is a consistent estimator of b_1, Zb_1 may be substituted with $Z\hat{b}_1$, the vector of fitted values from M_1. The t-test of α in:

$$[2.21] \qquad y = (1-\alpha)Xb_0 + \alpha Z\hat{b}_1 + v$$

is then the J-test of H_0: M_0 against H_1: M_1. The J test is not an exact test but it is only asymptotically valid and distributed as a standard normal variate. This is due to the fact that, under H_0, $(y-X\hat{b}_0)=\hat{e}_0$ is not independent of $Z\hat{b}_1$.

If Zb_1 is substituted by $y_{10}=Z(Z'Z)^{-1}Z'X(X'X)X'y$, the vector of fitted values from the regression of $X\hat{b}_0$ on Z, then the JA test is the test of $\alpha = 0$ in [2.21]. This is an exact test and it is

distributed as a t $(T-k_0-1)$ variate because y_{10} is independent of $\hat{e}_0$. However, when one of the models contains lagged dependent variables, then the JA test is only asymptotically valid and distributed as a standard normal variate.

The asymptotic equivalence of both the J and the JA test to the Cox test is due to the fact that the former can be directly derived from the Cox procedure, while the latter may be obtained from an asymptotic equivalent variation of the Cox test due to Atkinson (1970) (see McKinnon (1983)).

2.2 Parameter encompassing test

2.2.1 The F test

The F test for non nested models was derived in the context of the comprehensive model approach. In other terms, the non nested testing procedure is based on the construction of a general model obtained as the combination of two (or more) competing specifications. This general model may be given by some form of embedding, such as exponential embedding. In this way a parameter encompassing test of $H_0:M_0$ against $H_1:M_1$ is based on the F test of τ_1 in the model

$$[2.22] \qquad y = X\tau_0 + Z^{\sim}\tau_1 + w$$

where $Z^{\sim}$ and τ_1 are (Txr) and (rx1) matrices with r equal to the number of non overlapping variables in M_1. The F test is exactly distributed as $F(r,T-k_0-r)$ provided X and $Z^{\sim}$ are non stocastic matrices. In the case of dynamic models, where, for instance, lagged dependent variables are included among the regressors, the F test of τ_1 in [2.22] is only asymptotically valid and distributed as χ^2 (r,0) under H_0. It is worth observing that when Z contains only one column, that is only one variable, the F-test of τ_1 in [2.22] coincides with the J-test of α in [2.21]. However, when Z is characterized by more than one variable the two test

procedures will differ.

There are some well known problems with the application of the F-test: collinearity among the regressors in the matrices X and $Z^{\sim}$; the restricted number of degrees of freedom if X and $Z^{\sim}$ have great dimensions; various difficulties in case of non linearity. Besides this "technical" problems there is the difficulty of deciding what to do when both the F-test of τ_1 and τ_0 are significant. In effect, this would be interpreted as evidence in favour of the general model even if this has no economic meaning.

A new interpretation of F test is provided by Mizon-Richard (1986) who prove its asymptotic equivalence with n_2 in [2.3] (actually, Mizon-Richard show that n_2 and the F test adequately transformed coincide for all sample sizes). In this way, the F test for non nested models may be included among the class of parameter encompassing tests.

3. LOCAL POWER COMPARISONS

All four procedures we have illustrated in the previous section share the same asymptotic properties for the test of H_0:M_0 against H_1:M_1. In particular, as Davidson-McKinnon (1981) show; all tests have correct size asymptotically and they are all consistent in the sense that, as the sample size increases ($T\to\infty$) the probability tends to zero that these tests will fail to reject H_0 if a fixed H_1 is true.

Given the asymptotically equivalence of the tests, a possible comparison among them depends on the local behaviour of their power function. Two approaches may be followed according to the kind of alternative hypothesis being considered. In other words, we may check the power of different tests against fixed alternatives or a sequence of local alternatives. In the second case, the main motivation is to avoid trivial asymptotic distributions.

In general, a sequence of local alternative hypotheses, H_{1T},

approaches the null hypothesis as T tends to infinity, or, in other terms, the sequence differs from H_0 only by $O_p(T^{-\frac{1}{2}})$ terms (a somewhat different approach is analyzed in Davidson-McKinnon (1982) where the authors keep H_1 fixed and let the DGP approach H_0 so that it is possible to analyze the case where neither M_0 nor M_1 need be assumed true).

Various specifications for H_{1T} are possible. Dastoor-McAleer (1985) consider:

$$[3.1] \qquad H_G : y = Xb_0 + (T^{-\frac{1}{2}})\delta + \varepsilon$$

where δ is a vector of constants. As can be seen, H_G approaches H_0 as T tends to infinity. This form of general alternative is possible only when H_0 and H_1 are partially non nested in the sense given by McAleer-Pesaran (1985). As [3.1] is not informative about the rate at which the specific alternative H_1 approaches H_0, Pesaran (1982a) presents the following specification for H_{1T}:

$$[3.2] \qquad H_P : Z = XB + (T^{-\frac{1}{2}})\Gamma + o(T^{-\frac{1}{2}})$$

where B and Γ are $(k_0 x k_1)$ and $(T x k_1)$ matrices of constants (H_P may be shown to be a special case of H_G; see Dastoor-McAleer (1985)). It is interesting to observe that in the case of [3.2] the two matrices Z and X become more similar as T increases. However, if we want to keep the power bounded away from unity, the crucial condition is (see below):

$$[3.3] \qquad \lim_{T\to\infty} (\Gamma' A_0 \Gamma)/T = W_1$$

where $A_0 = (I - X(X'X)^{-1}X')$ and W_1 is a $(k_1 x k_1)$ matrix of constants. This occurs if the number of regressors of the model assumed as the null hypothesis is greater/equal the number of regressors of the alternative, or

$$[3.4] \qquad k_1 \geq k_0$$

(see Pesaran (1982a), page 1291).

By using [3.2], Pesaran derives the asymptotic power of the F, N_0^2, and J^2 test. The most important result is that all three testing procedures possess a limiting distribution with the same non centrality parameter (n) but with a different number of degrees of freedom. In particular:

$$\begin{aligned} F &\sim \chi^2(r,n) \\ J^2, N^2 &\sim \chi^2(1,n) \end{aligned} \qquad [3.5]$$

with $n = b_{11}' W_1 b_{11} / \sigma_1^2$

where b_{11} is the vector of coefficients of the non overlapping variables in the alternative model. As it can be seen from the formula for the non centrality parameter, condition [3.3] ensures that the value of n does not attain high values so that the power of the test - which is an incresing function of the non centrality parameter - is kept away from unity. Under [3.3] or [3.4], the asymptotic power is:

$$P_F = \lim_{T\to\infty} \text{prob}\{F \geq \chi_c^2(r,0) | H_{1T}\} \qquad [3.6]$$

$$= \text{prob}[\chi^2(r,n) \geq \chi_c^2(r,0)]$$

$$P_J = P_N = \lim_{T\to\infty} \text{prob}\{N_0^2 \geq \chi_c^2(1,0) | H_{1T}\}$$

$$= \text{prob}[\chi^2(1,n) \geq \chi_c^2(1,0)]$$

where P_F, P_J, P_N are the asymptotic power of the F, J_0^2, N_0^2, tests; c is the critical level of the test. By using a well established result in statistics (see Das Gupta-Perlman (1974)), Pesaran concludes that:

$$P_J = P_N > P_F$$

because of the difference in the number of degrees of freedom of the limiting distributions. Thus the asymptotic power of the F test is less than the asymptotic power of the Cox-type tests, the greater the number of non overlapping variables.

A serious limitation of Pesaran's approach is given by condition [3.4]. More general results may be obtained by the means of H_G. In this case, Dastoor-McAleer (1985) show that the local asymptotic power of the Cox-type tests and the F test can not be ranked, the reason being that not only the number of degrees of freedom of the non central F test is greater, but the non centrality parameter is greater as well.

Ericsson (1983) adopts a different form for the sequence of local alternatives. The general comprehensive model in [2.22]:

$$H_2 : y = X\delta_0 + Z\tilde{}\delta_1 + \varepsilon$$

may be considered as a local alternative hypothesis to H_0 in the particular case where

[3.7] $$(\delta_0,\delta_1) = (b_0,b_1) = (0,\delta/\sqrt{T})$$

The difference in [3.7] and in [3.2] is that in the former the parameters of H_0 and H_1 become more similar as $T\to\infty$. The advantage in this specification is in that it is not limited by any restrictions such as [3.4] and can be applied to the general case. In this context, Ericsson (1983) derives the asymptotic distribution of the N_0^2, and the F tests under the alternative. They are $\chi^2(1,n_c)$ and $\chi^2(r,n_F)$ where n_N, n_F are given by:

[3.8] $$n_c = \frac{\mu_0^2}{\omega_0}$$

with $$\mu_0^2 = -(\varepsilon_0+\varepsilon_1)/2\sigma_1\sqrt{\varepsilon_2}$$

$$\omega_0 = [(\varepsilon_0+\varepsilon_1)^2/4\varepsilon_2] \cdot VAR(q)/\sigma_1^2$$

$$\text{VAR}(q)/\sigma_1^2 = 4(\varepsilon_0+\varepsilon_1-\varepsilon_2)/(\varepsilon_0+\varepsilon_1)^2 - 4(\varepsilon_2-\varepsilon_3)/[(\varepsilon_0+\varepsilon_1)\varepsilon_2] + (\varepsilon_2-\varepsilon_3)/\varepsilon_2^2$$

$$\varepsilon_i = \phi_i - \phi_{i+1} \quad (i \geq 0)$$

$$\phi_0 = \text{plim}_1 \delta^{*\prime} Z^{\sim\prime} Z^{\sim} \delta^*/T$$

$$\phi_i = \text{plim}_1 \delta^{*\prime} Z^{\sim}[X(X'X)^{-1}Z^{\sim\prime}(Z^{\sim}Z^{\sim\prime})^{-1}Z^{\sim}]^{i-1}Z^{\sim\prime}(Z^{\sim}Z^{\sim\prime})^{-1}Z^{\sim}X\delta^*/T$$

$$\delta^* = b_{11}\sqrt{T}$$

$$n_F = \delta^{*\prime} Z^{\sim}[I-X(X'X)^{-1}X']Z^{\sim}\delta^*/(\sigma_1^2 \cdot T)$$

4. SMALL SAMPLE PERFORMANCE

While the F and the JA tests are exact tests - i.e. their exact distribution under the null is known - the Cox test and the J test are only asymptotically valid so that their finite sample behaviour may be assessed only by Monte Carlo analyses. However, when the models under consideration contain lagged dependent variables, then all testing procedures are only asymptotically valid and Monte Carlo techniques have to be employed for the F and the JA tests as well (see GP, page 149).

The Monte Carlo simulation studies that have been performed in the case of non nested linear regression models are well summarized by McAleer (1984). By considering the analyses of Pesaran (1974), Pesaran (1982a), Davidson-McKinnon (1982), and GP, McAleer provides a classification of the experiments according to different criteria, such as: the DGP, the relationship of the models to each other and also to the DGP when both models are false; the use of fixed or local alternatives; the sizes of the samples; the use of asymptotic or empirical critical values; the use of two-sided or one-sided tests; the importance of using symmetric or asymmetric empirical critical values; the robustness of the test statistics to non-normality of the disturbances; and the variety of tests examined (see McAleer (1984), page 53).

An established result is that the Cox test seems to reject the true null hypothesis too frequently, or, in other words, the estimated size is greater than the nominal size. If we want to perform an anlysis of the power of the test, then it is necessary to introduce some corrections which may concern the test statistic itself or the critical value employed in estimating the power of the test (for an application of this technique, see Ericsson (1986)). In the first case, interesting results are obtained by GP, who derive two different adjusted Cox tests.

The basic idea underlying the work of GP is that as T_0 in [2.8] has mean zero only asymptotically, a different specification should be found for which T_0 has mean zero in small samples as well. Formally, if we consider the numerator of NL_0 in [2.14]:

$$[4.1] \qquad z_0 = T/2(\hat{\sigma}_1^2-\hat{\sigma}_{10}^2)$$

a possible correction may be introduced by using its mean, that is:

$$[4.2] \qquad E(z_0) = \tilde{\sigma}_0^2\{Tr(R_0R_1)-k_1\}$$

where $R_0 = [X(X'X)^{-1}X']$ and $R_1 = [Z(Z'Z)^{-1}Z']$ (see GP, page 136).

In this way, the variate:

$$[4.3] \qquad \tilde{z}_0 = z_0 - \tilde{\sigma}_0^2[Tr(R_0R_1)-k_1]$$

has zero mean both asymptotically and in small samples ($\tilde{\sigma}_1^2$ is the unbiased estimator of σ_1^2, i = 0,1). following the analysis of Pesaran (1974), GP derive:

$$[4.4] \qquad W = \frac{(T-k_1)(\tilde{\sigma}_1^2-\tilde{\sigma}_{10}^2)}{(2\tilde{\sigma}_0^4Tr(B^2)+4\tilde{\sigma}_0^2e'_{100}e_{100})^{\frac{1}{2}}}$$

and

$$[4.5] \qquad \tilde{N} = \frac{\frac{1}{2}(T-k_1)\log(\tilde{\sigma}_1 - \tilde{\sigma}_{10})}{[(\tilde{\sigma}_0^2/\tilde{\sigma}_{10}^2)(e_{100}'e_{100} + \frac{1}{2}\tilde{\sigma}_0^2 Tr(B^2)]^{\frac{1}{2}}}$$

where $\tilde{\sigma}_i^2 = (e_i'e_i)/(T-k_i) \qquad (i=0,1)$

$$\tilde{\sigma}_{10}^2 = \{\tilde{\sigma}_0^2[Tr(A_0A_1)] + (e_{10}'e_{10})\}/(T-k_1)$$

$$Tr(B^2) = k_1 - Tr(R_1R_0)^2 - [k_1 - Tr(R_1R_0)]^2/(T-k_0)$$

The small sample corrections to the basic Cox test do not affect its asymptotic distribution so that W and $\tilde{N}$ possess the same asymptotic standard normal distribution and the W^2 and the $\tilde{N}^2$ the asymptotic $\chi^2(1,0)$ under the null hypothesis. When W^2 and $\tilde{N}^2$ are computed under the alternative, then the asymptotic distribution of the two procedures is not affected (Ericsson (1983) formally proves that a slightly different form of $\tilde{N}^2$ is asymptotically equivalent to N_0 under H_1).

For finite samples, however, no theoretical results are available so only Monte Carlo simulations may shed light on the behaviour of W^2 and $\tilde{N}^2$. In a series of different experiments (equal/unequal number of regressors in the two competing models; non normal errors; presence of a lagged dependent variable) GP find that the mean adjusted Cox tests (W^2 and $\tilde{N}^2$) have empirical sizes that are very close to the nominal ones. In particular, for the set of experiments where a lagged dependent variable is included among the regressors (the case in which we are mainly interested), the performance of W^2 is very good. Indeed, for a sample size of 20, GP find that $\tilde{N}^2$ and J have too high size while W, JA and F present values that are not very far from the nominal ones. The most interesting result concerns the behaviour of the W test in terms of power. In effect, the power estimate of W is the highest among the testing procedures under examination.

As to the JA test while it shows an acceptable empirical size, its power performance seems to be affected by the number of regressors in the two competing specifications. In particular,

when the model which is supposed to be the null hypothesis contains more regressors than the alternative, the power estimate of the JA test is very low.

The J test exhibits the same problem as the Cox test, that is its empirical size is always higher than the nominal one. An adjustment similar to that proposed for the Cox test may be applied but in this case the main motivation for the computation of the J test, that is its simplicity, is lost.

5. THE EXPERIMENTAL DESIGN

In order to apply a post simulation analysis technique to the results of GP, we replicate the experiment of GP in the case of the presence of a lagged dependent variable in both competing models. This experiment, or, better, set of experiments (labelled as "D" in GP, page 142) is interesting also because it considers the case of a different number of regressors in the two specifications.

The data generation process adopted by GP is:

$$[5.1] \qquad y_t = \beta y_{t-1} + \sum_{i=1}^{k_0} \alpha_i x_{it} + u_{0T}$$

with $u_{0t} \sim NID(0,\sigma_u^2)$ and $x_{it} \sim NID(0,1)$. x_I and u_i are generated by the means of the routine G05DDF of the FORTRAN 77 NAGLIB. Setting $\alpha_i=1$ for every i (see below), the variance of u_{0t} is calculated in such a way that the value of the population multiple correlation coefficient is R^2. So:

$$[5.2] \qquad \sigma_0^2 = k_0(1-R^2)/(R^2-\beta^2) \qquad \beta^2 < R^2 < 1$$

The presence of a lagged dependent variable creates some problems for the generation of initial values. Two solutions have been proposed in the literature. The "conventional" method is based upon the generation of y_t with y_0 set equal to some arbitrary

value. The sample of y_t is then obtained by discarding a certain number of initial values. The alternative method is much more concerned with the first observation, y_0. This is generated randomly and such that the variance is equal to the variance of y_t, as specified in [5.2]. In our case, given [5.1] and [5.2], y_0 is generated as a NID$(0,\sigma_{y0}^2)$ where:

$$[5.3] \qquad \sigma_{y0}^2 = k_0/(R^2-\beta^2)$$

The second method is employed by GP beacuse, as they say: "it seems superior on theoretical and computational grounds" (GP, page 141).

The alternative model is given by:

$$[5.4] \qquad y_t = \beta^* y_{t-1} + \sum_{i-1}^{k_1} z_{it} + u_{it}$$

where z_{it} is obtained as:

$$z_{it} = \mu_i x_{it} + v_{it} \qquad \text{for } i = 1,2,\ldots,\min(k_0,k_1)$$

[5.5] and if $k_1 > k_0$

$$z_{it} = v_{it} \qquad \text{for } i = k_0+1,\ldots,k_1$$

v_{it} ~ NID(0,1) and is generated with the same subroutine as the x's and the u's. The coefficient μ_i is determined by:

$$[5.6] \qquad \mu_i = \tau_i/(1-\tau_i^2)^{\frac{1}{2}}$$

where τ_i is the canonical correlation between x_i and z_i.

GP calculate the tests: N_0^2, $\tilde{N}^2$, W^2, J^2 and F under the null hypothesis for 500 replications of each of 144 points defined by the combination of:

$$
\begin{aligned}
& T = (20, 40, 60) \\
& R^2 = (0.5, 0.8) \\
[5.7] \qquad & \tau^2 = (0.3, 0.9) \\
& \beta = (0.3, 0.5, 0.7) \\
& u_t \sim NID(0, \sigma_u^2)
\end{aligned}
$$

In our case, however, we have decided to reduce the number of experiments by ignoring the cases given by $\beta = 0.5$ and $T = 40$ so that the number of our experiments is 64.

In this context, the Monte Carlo design variables are (see Hendry (1984) and Ericsson (1986)):

$$[5.8] \qquad q = (\beta, \alpha_i, \tau_i^2, R^2, k_0, k_1) \in Q = \{q | \tau_i^2 < 1,\ \beta^2 < R^2 < 1\}$$

$$T \in T = [\ 20, 60\ ]$$

In this way,

$$SP = (Q \times T)$$

defines the parameter space.

The objective of GP's Monte Carlo simulation is to investigate the finite sample rejection frequency:

$$[5.9] \qquad \pi = \text{prob} (|\theta| \geq c)$$

where θ is any of the N_0^2, $\tilde{N}^2$, W^2, J^2, JA^2, F tests and c the nominal critical value (generally correspondent to the 5% significance level).
π depends upon q and T and it may be interpreted as a conditional probability:

$$[5.10] \qquad \pi = \text{prob} \{ |\theta| \geq c \,|q,T \} = G^* (q,T)$$

It has to be noticed that c is a design veriable and as such it ought to be included in q. However, as it is kept constant in all experiments, its presence in q can be ignored. Further, in order to simplify the calculations, the coefficients of x_i, α_i, are set equal to 1 and $\tau_i^2=\tau^2$ (see Pesaran (1982a) and GP).

All estimations are computed in deviations from the mean in order to take into account the presence of a constant term. In this way the mean of the estimated residuals is zero and this is perfectly coherent with the generation of the structural disturbances as $N(0,\sigma_u^2)$. In the analysis of GP this does not happen because they estimate a model without a constant term but generating u as $N(0,\sigma_u^2)$. Clearly, their estimate of the variance of u is biased and may cause trouble in the derivation of the test procedures under examination.

The generation of the alternative set of regressors according to [5.5] and [5.6] highlights the fact that the relevant control variable is given by the correlation between x_i and z_i. Clearly, as τ^2 increases the distance between the two models decreases. In this context, it is interesting to derive the non centrality parameter characterizing the asymptotic distributions of the testing procedures under the alternative. When the number of regressors in both models is equal - i.e. $k_0 = k_1$ - we may apply Pesaran's (1982a) analysis and derive the non centrality parameter according to [3.5]. In this case, the sequence of alternative hypotheses is given by [3.2] and a measure of the distance between the two models may be:

$$[5.11] \qquad \delta = T(1 - \tau^2) \qquad (\delta_i = \delta)$$

The non centrality parameter is then obtained from [3.5] as:

$$[5.12] \qquad n = \delta(R^2-\beta^2)/(1-R^2)$$

When the two models are charaterized by a different number of regressors, then we may apply Ericcson's (1986) formulae for the non centrality parameters as given in [3.8]. In this context, we

obtain:

A) $k_1 \geq k_0$

[5.13]
$$n_c = \frac{T \cdot k_0[(1-\tau^2)(1+\tau^2)]\tau^2(R^2-\beta^2)}{[\tau^2 k_0+(1-\tau^2)k_1][4\tau^2+(1-\tau^4)(1+\tau^2)](1-R^2)}$$

$$n_F = \frac{T \cdot k_0(1-\tau^2)(R^2-\beta^2)}{[\tau^2 k_0+(1-\tau^2)k_1](1-R^2)}$$

B) $k_0 \geq k_1$

[5.14]
$$n_c = \frac{T \cdot k_0[(1-\tau^2)(1+\tau^2)^2]\tau^2(R^2-\beta^2)}{k_1[4\tau^4+(1-\tau^4)(1+\tau^2)](1-R^2)}$$

$$n_F = \frac{T \cdot k_0(1-\tau^2)(R^2-\beta^2)}{k_1(1-R^2)}$$

It is interesting to observe that if we increase k_0, then n_c and n_F increase as well. This result may be explained by the way we have generated the variables x and z (see [5.5] and [5.6]). In effect, in case A), given the DGP:

[5.15] $$y_t = \beta y_{t-1} + x_1 + \varepsilon$$

and the alternative:

[5.16] $$y_t = \beta^* y_{t-1} + z_1 + z_2 + w$$

using [5.5] we have:

[5.17] $$y_t = \beta^* y_{t-1} + \tau x_1 + v_1 + v_2 + w$$

If we increase k_0, then the two models become:

[5.18] $$M_0 : y_t = \beta y_{t-1} + x_1 + x_2 + \varepsilon$$

$$M_1 : y_t = \beta^* y_{t-1} + \tau(x_1+x_2) + v_1 + v_2 + w$$

and the distance between them is greater.

When $k_0 = k_1$, [5.13] and [5.14] reduce to the formulae given by Ericsson (1986) (page 6).

Ericsson observes that if $\tau^2 \simeq 1$, then Pesaran's result - i.e. $n_c = n_F$ - obtains. In effect, Ericsson's approach and Pesaran's are very similar numerically. In this context, when $k_0 = k_1$, we apply Pesaran's formula for the non centrality parameter as given in [5.12].

6. MONTE CARLO METHODOLOGY

In this section, we explain the basic methodology underlying our exercise. Particular attention is devoted to the concept of estimated size and power in order to obtain two estimates which can be considered homogenous.

The basic priciples characterizing our response surface analysis are then illustrated, along with the "battery" of misspecification tests to check the results of the estimation procedure.

6.1 Size and power estimates

As indicated in the previous section, the objective of our Monte Carlo simulation is the investigation of the small sample behaviour of different non nested testing procedures. The small sample performance of a test may be measured by its estimated size and power. The general concept underlying the estimation of these two quantities is represented by expression [5.10]. In particular, the estimated size is computed as the proportion of times the test

rejects the null hypothesis when it is in fact true. As it may be seen, the estimated size is an estimate of the probability of type I error. The nominal value is given by the 5% critical value of:

- the $\chi^2(1,0)$ distribution for the N_0^2, $\tilde{N}^2$, W^2, J^2, JA^2 tests (under H_0).

In the case of the F test, we follow GP who computes the estimates of the size and the power by using the nominal critical value of the F distribution. It is worth pointing out that this choice may be justified only on practical grounds, since in the case of dynamic models the F test is asymptotically distributed as a χ^2.

The estimate of the power is obtained by interchanging the role of the null and the alternative hypotheses. In this way, an estimate of the power is the proportion of times the test rejects the null hypothesis.

In the context of non nested testing procedures, it is important to realize which kind of power and size estimates are computed. Indeed there are four different cases according to which of the two models is assumed to be the DGP. By computing the proportions of times the test rejects the null hypothesis, we have the following two cases:

1) the DGP is M_0 (see [5.11])

1a) $H_0 : M_0$ and $H_1 : M_1$ estimate of the size of the test of M_0;

1b) $H_0 : M_1$ and $H_1 : M_0$ estimate of the power of the test of M_1;

2) The DGP is M_1 (see [5.14])

2a) $H_0 : M_0$ and $H_1 : M_1$ estimate of the power of the test of M_0;

2b) $H_0 : M_1$ and $H_1 : M_0$ estimate of the size of the test of M_1;

It is worth observing that the results of the two sets of tests (1 and 2) are the same when the models are perfectly symmetric. A case in which this may happen is when we have the same number of non overlapping variables in both competing specifications (see Ericsson (1986), page 23). In this case, it is

more convenient to compute the size and power estimates by considering only one set of tests (in our case: set 1). If, on the other hand, there is some form of asymmetry between the two models, such as a different numbers of regressors, then it may be the case that the results of the two sets of tests are asymmetric as well. In this context, both case 1 and case 2 have to be analyzed and this implies that we have to interchange the DGP. When we use [5.4] as the DGP, the variables z_i generated from the first of [5.5] are scaled so that the variance of y_t and u_t is not changed: in this case, the scaling factor is:

$$Sc = [\ (1+\mu^2)\ ]^{-\frac{1}{2}}$$

6.2 Response surface analysis

The main drawbacks of every Monte Carlo simulation may be summarized in two points: imprecision and specificity.

a) The imprecision of results.

It is well established that when performing simulations the results are always imprecise even when the sample size is large (see Mizon-Hendry (1980)). A measure of the imprecision is provided by the standard error of the size/power estimate. This is computed as:

$$[6.1] \qquad Ser(\pi) = \sqrt{[\pi(1-\pi)/N]}$$

where π is an estimate of π given by [5.10]; N is the total number of replications.

Imprecision may be reduced by the so called variance reduction techniques (see Hendry (1984)). In our case, we have applied one of these techniques, namely the use of the same sequence of random numbers across different experiments. In this way, the re-using of the same set of random numbers may help in reducing the "intra-experiments" variability of the estimates (for a brief exposition of other techniques, see Hendry (1984)).

b) The specificity of results.

Every experiment depends upon the value of the design variables. In effect, if P is the small sample power function of a test, then:

$$[6.2] \qquad P = G^{*}(q,T)$$

(see section 5).

The main interest of a researcher is not to obtain estimates of P related to a few points within SP = q x T, the parameter space, but to have a more general view of the behaviour of P across SP. To achieve these results, the Monte Carlo estimate of P, $\hat{P}$, may be used but as Mizon-Hendry (1980) show, it is very inaccurate. In effect, by considering its standard error (given by [6.1]) N, the number of replications in each experiment should be very high in order to obtain an acceptable estimate. An efficient alternative is offered by the response surface analysis where $\hat{P}$, the response is interpreted as a function of the parameters characterizing the Monte Carlo experiment.

In general terms, we have:

$$[6.3] \qquad \hat{P} = G(q,T) + e$$

where G (q,T), the response surface, is an approximation to G^{*} (q,T) and e is a disturbance term. In particular, it is worth observing that e is given by two components: a simulation component related to the use of $\hat{P}$ instead of P; and an approximation component connected to the use of G (q,T) instead of G^{*} (q,T). A requisite of [6.3] is that the form of G should ensure that all power estimates and predictions lie within the (0,1) interval. Following Mizon-Hendry (1980) and Ericsson (1986) a possible response surface may be:

$$[6.4] \qquad \frac{\hat{P}}{(1-\hat{P})} = \frac{\left(P_A\right)^{j}}{1-P_A} \cdot \exp[G(q,T)]$$

where the use of P_A, the asymptotic power of the test, may increase the efficiency of the analysis.

In order to derive an implementable form of [6.4] we may resort to some results obtained by Cox (1970) about bynary variables (see Ericsson (1986), pages 6-8). In effect, if $\hat{P}$ is regarded as the probability of success (rejection of the false model) with $0 < \hat{P} < 1$ and $N > 1$ - then letting:

[6.5] $A = [S(N-S)]/(N-1)$ where S is the number of replications characterized by success (in our case, S=NP)

[6.6] $L(h) = A^{\frac{1}{2}}\log[h/(1-h)]$ with $0 < h < 1$ where h is the probability of a binary variable

[6.7] $L^*(h) = A^{\frac{1}{2}}\log[(h-d)/(1-h-d)]$ $d < h < 1-d$ where $d=(2N)^{-1}$

then Cox (1970) shows that:

[6.8] $$\varrho\,(\hat{P},P) \simeq L^*(\hat{P})-L(P) \sim N(0,1)$$

As P is unknown, it is important to check if the asymptotic power of the test, P_A, equals P and to do this we may simply check whether $\varrho\,(\hat{P},P_A)$ is significantly different from the standard normal distribution. The asymptotic power is computed by applying formulae [3.6] with appropriate non centrality parameters. The problem of approximating a non central χ^2 distribution is analyzed by Mizon-Hendry (1980), Hendry (1984), Ericsson (1986). The method we use is based on Mizon-Hendry (1980) or the first one described by Ericsson (1986) in his Appendix C.

Equations [6.4] and [6.8] constitute the basis for our response surface analysis. In effect, after some manipulations is possible to obtain:

[6.9] $$L^*(\hat{P}) = jL(P_A) + A^{\frac{1}{2}}G(q,T) + \varepsilon$$

where ε is distributed as NID(0,1).

In the response surface given by [6.9], j and G(q,T) are unknown and some approximations have to be employed. In particular, j may be expanded in powers of $T^{-\frac{1}{2}}$ around $T = \infty$, so that we have:

$$[6.10] \qquad j = j_0 + j_1 T^{-\frac{1}{2}} + \ldots$$

Things get more complicated for G(.,.). G (.,.) may be expanded in powers of $T^{-\frac{1}{2}}$ and of the elements of q. In this last case, the expansion implies the computation of the powers and cross-products of the elements of q. For instance, as we have five elements in q, if we compute the second-order Taylor approximation of G($q,T^{-\frac{1}{2}}$) in terms of q we have to take into account 15 terms.

By truncating the expansion of j and G($q,T^{-\frac{1}{2}}$) at a suitable order, that is when the approximation error is negligible, estimates of the coefficients of j and G($q,T^{-\frac{1}{2}}$) may be obtained by the application of the least squares to:

$$[6.11] \qquad L^{*}(\hat{P}) = j_0 L(P_A) + j_1 T^{-\frac{1}{2}} L(P_A) + \ldots + A^{\frac{1}{2}} T^{-\frac{1}{2}} H(q,T^{-\frac{1}{2}}) + \varepsilon$$

where H is the least squares approximation to G. The presence of $A^{\frac{1}{2}}$ is needed is we want to correct for heteroscedasticity.

The usefulness of a response surface as the one shown in [6.9] lies in the fact that it can offer a valid synthesis of the bulk of Monte Carlo results. More importantly, response surfaces are of great interest because they help predicting the values of P for different points in Q not included in the simulation. In this way an acceptable response surface can be used as an approximation to the finite sample distribution function of the statistic under examination. At this stage, the problem is then to establish the validity of the response surface employed.

Criteria to check the approximations for j and G (.,.) are provided by asymptotic theory and the fact that ε in [6.9] must be asymptotically distributed as NID(0,1).

In the first case, asymptotic theory implies that j_0 must not be significantly different from one. In the second case, the properties of ε may be checked by a set of misspecification tests

well established in the econometric practice. Practically, we want to check for independence, homoscedasticity, normality and stability of the estimated residuals. In effect, when H(.,.) = G(.,.), then the estimated residuals must be distributed as NID(0,1). A list of a possible tests is provided by Ericsson (1986, page 11a). We apply almost the same set of tests, except for the normality test and the test for heteroscedasticity in quadratic terms. Specifically, the normality test is substituted with the set of tests proposed by Kiefer-Salmon (1983). The reason for this choice is that with the Kiefer-Salmon specification we can test not only for skewness and kurtosis but for higher moments of the ε distribution as well. The whole set of misspecification tests is summarized in table 1. The use of tests to detect the presence of autocorrelation deserves some comments. In effect, our analysis lies in a cross-sectional set-up so that these tests may appear meaningless. However, a good response surface must be characterized by an error term which is independent across experiments. This must hold for any ordering of experiment. Once an ordering of the experiments is chosen, then the possibility of autocorrelation (due for instance to incorrect specification of H(.,.)) has to be checked. As a final remark, it should be noticed that all these tests help to check if a correct inference can be applied to the specification we have obtained.

7. RESULTS OF MONTE CARLO AND RESPONSE SURFACE ANALYSIS

The purpose of our Monte Carlo simulation is to replicate one of the sets of experiments of GP, mainly the set labelled as "D" (see GP, page 142). This set of experiments, as we have shown in section 5, is of great interest because it tackles the problem of the small sample behaviour of non nested testing procedures in the case of dynamic models. Moreover, the case of different number of regressors in the competing models is considered.

The results of the Monte Carlo simulation for the W^2 and F tests - i.e. the power estimates of these tests - are then used in

Table 1

Misspecification Tests

Misspecification	Statistic	Sources
$j_0 = 1$	$\mu_1(1,E-v-1)$	Hendry (1984, p. 962)
p invalid parameter restrictions	$\mu_2(p,E-v-p)$	Johnston (1963, p. 126)
Skewness (SK)	$\tau_2(1)$	Kiefer-Salmon (1982)
Excess kurtosis (EK)	$\tau_3(1)$	"
SK + EK	$\tau_4(2)$	"
SK + EK + 5^{th} moment	$\tau_5(3)$	"
First-order ARCH	$\tau_6(1)$	Engle (1982)
First-order residual DW autocorrelation		Durbin-Watson (1950-51)
	$\mu_3(p,E-v-p)$	Harvey (1981, p. 173)
	$\tau_7(p)$	Box-Pierce (1978)
$H^*(...)$ non costant over subsamples	$\mu_4(v,E-2v)$	Kendall (1946, p. 242 ff.)
	$\mu_5(E_1-vn\ E_2-v)$	Chow (1960, p. 595 ff.)
Predicted failure over a subset of p observations	$\mu_6(p,E-v-p)$	Chow (1960, p. 594-5)
	$\tau_8(p)$	Hendry (1979, p. 222)

Notes: $\tau_i(s)$ represents a χ^2 distributed test statistic with s degrees of freedom;
$\mu_i(a,b)$ represents a F-distributed test statistic with a,b degrees of freedom.
E is the number of experiments or observations.

the response surface analysis presented in paragraph 7.2. The outcome of this analysis is very useful because it may give a synthetic description of the Monte Carlo simulation without having to analyze a huge quantity of tables and numbers.

7.1 Simulation results

The results the of our simulations are numerically different from GP study, although qualitatively they show the same pattern[1]. There are various reasons for such a difference. In particular, as explained in section 5, all our estimations are characterized by the presence of an intercept term in both models, while GP's analysis do not include any constants.

Concerning the J^2 and the JA^2 tests, another source of difference is given by the nominal critical value we use. As GP clearly show, the have derived the estimated size and (power) for the J^2 and the JA^2 tests using the nominal critical value from the $F\ (1,\ T-k_i-2)$ $(i = 1,0)$ distribution, whereas in our case, the nominal critical value is obtained from the $\chi^2\ (1,0)$ distribution. In effect, all our estimates for the J^2 and the JA^2 tests are greater than GP's ones and this agrees with the fact that:

$$F\ (1,\ T-k_i-2) > \chi^2\ (1,0)$$

for $T = 20,\ 60$.

Another important observation concerns the problem of the kind of power we actually compute. In order to derive the power of the test of M_0, we have to carry out the set of computations under case 2a) illustrated in section 6.

Following Ericsson (1986), when the models are symmetric, the extra computations under 2a) are not necessary because the

---

1. Tables for the experiments analyzed with Monte Carlo simulations are avalaible from the author on request.

results are the same as 1b). This is no longer true when the number of non overlapping regressors in both models is different. In our case, however, extra computations may be avoided by observing that the experiments characterized by:

a) DGP : M_0, $(k_0, k_1) = (4, 2)$

and

b) DGP : M_1, $(k_0, k_1) = (2, 4)$

may be considered as symmetric. This seems, in effect, supported by the results we have obtained by analyzing the two experiments:

E1)		and E2)	
	$R^2 = 0.5$		$R^2 = 0.5$
	$\tau^2 = 0.3$		$\tau^2 = 0.3$
	$\beta = 0.3$		$\beta = 0.3$
	$T = 20$		$T = 20$
	$(k_0,k_1) = (2,4)$		$(k_0,k_1) = (4,2)$

under different DGP's.

If we compare the results (see table 2) of the two experiments when the GDP is given by M_0 and M_1, then it is possible to realize that the estimates for E2) when DGP : M_1 may be approximated by the estimates of E1) when DGP : M_0 (and viceversa). If this finding reveals correct it allows to avoid extra computations in cases where there is asymmetry between the competing models. Clearly, more work, both theoretical and empirical is needed. Our estimates are related to the test of M_0 while, from what we may understand from the tables presented by GP, their power estimates have to be considered as related to the test of M_1 (case 1b) at page 30). As a consequence, the outcome about the power of the JA test is completely reversed in our case, because if the power of the test of M_0 is computed, then the JA test tends to show a very poor performance when $k_0 < k_1$. On the contrary, GP

Table 2

Results of experiments E1 and E2

Exp. E1	DGP : M_0				DGP : M_1			
Test	Est. size		Est. power		Est. size		Est. power	
N^2_0	33.80	(2.12)	90.40	(1.32)	23.80	(1.90)	96.20	(0.86)
W^2	4.60	(0.94)	68.80	(2.07)	6.40	(1.09)	56.00	(2.22)
$\tilde{N}^2$	9.00	(1.28)	76.60	(1.89)	8.80	(1.27)	69.60	(2.06)
J^2	26.20	(1.97)	80.00	(1.79)	11.20	(1.41)	92.40	(1.19)
JA^2	7.60	(1.19)	61.60	(2.18)	6.40	(1.09)	31.60	(2.08)
F	4.20	(0.89)	58.40	(2.20)	3.80	(0.86)	54.80	(2.23)

Exp. E2	DGP : M_0				DGP : M_1			
Test	Est. size		Est. power		Est. size		Est. power	
N^2_0	23.00	(1.88)	99.90	(0.44)	37.60	(2.17)	92.20	(1.20)
W^2	4.80	(0.96)	63.00	(2.16)	4.60	(0.94)	65.60	(2.12)
$\tilde{N}^2$	6.80	(1.13)	75.40	(2.06)	8.40	(1.24)	78.00	(1.85)
J^2	10.00	(1.34)	97.00	(0.76)	26.80	(1.98)	79.40	(1.81)
JA^2	8.00	(1.21)	31.20	(2.07)	7.40	(1.17)	60.60	(2.19)
F	5.00	(0.97)	45.80	(2.23)	3.60	(0.83)	53.80	(2.23)

Note: The standard error is in parentheses.
All values are expressed as percentages.

obtain results supporting the conclusion that the small sample power of the JA test is very poor when $k_0 > k_1$ (see GP, page 150).

In terms of size estimates, the main result we can derive from our analysis is that the small sample correction for the Cox test, which allows to obtain the W^2 and the $\tilde{N}^2$ tests, seems to work in the desired way. This is mainly true for the W^2 test.

In order to check if the estimated power of the W^2 and the F tests significantly differs from the asymptotic power, we can perform a simple test by using the result derived by Ericsson and reproduced in [6.6]. Table 3 reports the value of the ratio between the mean of $\Omega(\hat{P},P_A)$ and the corresponding standard error (this ratio is represented by: $\Phi(\hat{P},P_A)$) for the following cases: the total number of experiments (GEN); the experiments characterized by T = 20 and T = 60.

Table 3

Case 1 : W^2

	GEN	T=20	T=60
$\Phi(\hat{P}, P_A)$	0.322	0.433	0.083

Case 2 : F

$\Phi(\hat{P}, P_A)$	-0.201	0.039	-0.061

As can be seen, in all cases the test does not reject the hypothesis that the small sample estimate of the power for the two tests under consideration is significantly different from the asymptotic power.

7.2 Response surface analysis

In this paragraph we apply the theoretical analysis presented in section 6. Taylor approximations for j and G(.,.) are derived and truncated at the second order. In effect, more com-

plicated response surfaces are not implementable, given the number of observations at our disposal. Specifically, the number of experiments used is 44 and 43 for the case of the W^2 and F tests respectively. Moreover, the data relative to 5 experiments have been left for post sample analysis, while 15 and 16 observations have been excluded. This is due to the fact these observations coincide with extreme values of P_A, that is for $P_A \simeq 99.999$ (see Mizon-Hendry (1980) and Hendry (1984)).

The estimates of the general regression equations for the two cases under consideration are illustrated in table 4.

At first blush, it appears that in the general specifications the fit of the response surfaces is not so bad; however, the estimates of the standard error of the regression are significantly different from one and this violates one of the basic assumptions of response surface analyses.

In the case of the coefficient of $L(P_A)$ our results are contrasting: the value of this coefficient is not significantly different from 1 for the F test, while it is when W^2 is concerned.

The general specifications of the response surfaces shown in table 4 are only the first step in the analysis. Indeed, they may be improved by dropping those variables whose estimated coefficients are not significant. It is importat to observe that this procedure, often applied in econometrics, is correct only if the results of the misspecification tests support the distributional assumptions about the response surface. In this context, the specification used in the case of the W^2 test is characterized by a distribution of the residuals which seems to be non normal - see the value of the normality tests. As a consequence the application of the aforementioned procedure is not theoretically correct, but we may apply it in order to see if it is possible to obtain more interesting results. The outcome of this operation is shown in table 5.

The two restricted response surfaces show a slight improvement in terms of the standard error estimates and the value of the misspecification tests (see table 6).

A first explanation of the results presented in the tables

4-5-6 may be found in the particular nature of the experiments we have carried out in our Monte Carlo simulations. These experiments concern primarily fixed alternatives (see GP, page 141), whereas response surface analyses of the kind we are considering would require the computation of power estimates under a sequence of local alternatives. A possible consequence of this fact is the relatively high number of extreme values excluded from the analysis.

Apart from these observations, some interesting remarks may nevertheless be made. In effect, the least squares approximation has been shown to possess some nice properties even when the model under consideration is misspecified (see White (1980)). In this way, it is interesting to observe that some of the results presented by GP can be easily confirmed. This concerns, for instance, the effect of the coefficient of the lagged dependent variable (β): as GP point out (GP, page 150) there is a negative relationship between the value of this coefficient and the power of the tests. In other terms, as β increases the two models tend to be less indistinguishable as confirmed by the value of the non centrality parameter given by the formulae [5.12], [5.13], [5.14]. The same reasoning applies for the correlation coefficient between the two sets of regressors (τ^2).

All these observations are straightforward by examing the response surface, while it should not be so if we had to go through all the numbers presented in the tables. This is one example of the advantages of response surface analyses. Moreover, if the response surface is well calibrated, then it could be interpreted as an approximation of the small sample power function of the test under consideration and could also be used to predict values outside the range of experiments analyzed by Monte Carlo techniques (in our case, the predictive failure tests are not so significant as it is for other tests; see, for instance, the value of Chow test and Hendry test - μ_6 and τ_8 in table 6 - especially for the F power estimates).

Table 4

General Specifications

Variable	F	(Ser)	[Cser]	W^2	(Ser)	[Cser]
$L(P_A)$	0.84	(0.29)	[0.13]	1.52	(0.32)	[0.13]
$L(P_A)/T^{1/2}$	-1.83	(1.51)	[0.69]			
$L(P_A)/T$						
$L(P_A)/T^{3/2}$				-112.67	(29.05)	[12.87]
$\beta\Leftrightarrow$	7.19	(12.57)	[6.57]	-49.01	(12.60)	[6.13]
$\beta^{**}2$	0.41	(0.79)	[0.48]	-0.53	(0.60)	[0.26]
τ^2	-0.12	(1.53)	[0.82]	-7.26	(1.56)	[0.84]
$\tau^{2**}2$	-0.03	(0.04)	[0.02]	0.04	(0.02)	[0.02]
R^2	0.62	(7.69)	[3.87]	91.34	(20.63)	[11.02]
$R^{2**}2$	-0.72	(0.58)	[0.36]	-4.33	(1.46)	[0.69]
K_0	-0.61	(1.25)	[0.63]	-2.14	(1.35)	[0.86]
$K_0^{**}2$	0.07	(0.02)	[0.01]	0.07	(0.02)	[0.008]
K_1	-0.19	(1.10)	[0.11]	0.39	(1.37)	[0.64]
$K_1^{**}2$	0.04	(0.01)	[0.01]	0.03	(0.02)	[0.008]
$\tau^{2*}\beta$	0.19	(0.19)	[0.11]	-0.34	(0.12)	[0.07]
$\tau^{2*}R^2$	-0.08	(0.07)	[0.09]	0.29	(0.21)	[0.13]
$\tau^{2*}K_0$	-0.05	(0.02)	[0.01]	-0.38	(0.02)	[0.01]
$\tau^{2*}K_1$	0.06	(0.09)	[0.01]	0.06	(0.01)	[0.008]
$R^{2*}\beta$	0.89	(0.68)	[0.38]	4.75	(1.00)	[0.68]
$R^{2*}K_0$	0.06	(0.09)	[0.07]	0.02	(0.20)	[0.10]
$R^{2*}K_1$	0.004	(0.09)	[0.07]	-0.12	(0.15)	[0.09]
β^*K_0	-0.17	(0.13)	[0.08]	-0.14	(0.11)	[0.07]
β^*K_1	0.08	(0.13)	[0.09]	0.06	(0.11)	[0.05]
$K_0^*K_1$	-0.07	(0.02)	[0.01]	-0.09	(0.02)	[0.008]
A	2.12	(2.41)	[1.39]			
$A^*T^{-1/2}$	-13.64	(12.99)	[2.03]	-13.59	(4.73)	[3.04]
R squared	0.94			0.94		
σ	3.73			4.01		
n. obs.	44.00			43.00		

Notes: $\Leftrightarrow$ all following variables, except for A and $A^*T^{-1/2}$, are multiplied by $A^*T^{-1/2}$.
Ser is the standard error of the estimated coefficient.
Cser is the standard error corrected for eteroschedasticity;
R^2, τ^2, β, K_0, K_1 are the parameters used for Monte Carlo analysis.

Table 5

Restricted Specifications

Variable	F	(Ser)	[Cser]	W^2	(Ser)	[Cser]
$L(P_A)$	0.84	(0.15)	[0.10]	1.64	(0.20)	[0.14]
$L(P_A)/T^{½}$	-1.63	(0.74)	[0.55]			
$L(P_A)/T$						
$L(P_A)/T^{3/2}$				-123.00	(19.09)	[14.09]
$\beta^{\Leftrightarrow}$				-51.30	(8.83)	[6.55]
β^{**2}						
τ^2				-7.19	(1.28)	[0.87]
τ^{2**2}				0.05	(0.02)	[0.02]
R^2				95.50	(14.09)	[11.80]
R^{2**2}				-4.53	(0.84)	[0.55]
K_o	-1.27	(0.48)	[0.36]	-1.66	(1.05)	[0.07]
K_o**2	0.07	(0.01)	[0.009]	0.07	(0.01)	[0.07]
K_1						
K_1**2				0.03	(0.009)	[0.004]
$\tau^2*\beta$	0.14	(0.05)	[0.04]	-0.32	(0.09)	[0.004]
τ^2*R^2	-0.14	(0.06)	[0.04]	0.29	(0.15)	[0.10]
τ^2*K_o				-0.04	(0.01)	[0.007]
τ^2*K_1						
$R^2*\beta$	0.38	(0.24)	[0.19]	4.32	(0.85)	[0.61]
R^2*K_o						
R^2*K_1						
$\beta*K_o$	-0.11	(0.07)	[0.05]	-0.14	(0.08)	[0.05]
$\beta*K_1$						
K_o*K_1	-0.05	(0.008)	[0.07]	-0.09	(0.02)	[0.008]
A	0.40	(0.33)	[0.32]			
$A*T^{-½}$				-15.02	(3.87)	[3.53]
R squared	0.94			0.93		
σ	3.49			3.46		
n. obs.	44.00			43.00		

Notes: $\Leftrightarrow$ all following variables, except for A and $A*T^{-½}$, are multiplied by $A*T^{-½}$.
Ser is the standard error of the estimated coefficient.
Cser is the standard error corrected for eteroschedasticity;
R^2, τ^2, β, K_o, K_1 are the parameters used for Monte Carlo analysis.

Table 6

Misspecification Tests

Statistic	W^2 gen.	F gen.	W^2 res.	F res.
j = 1	15.06*	1.53	10.15*	0.011
μ_2			0.35	0.33
τ_2	9.51*	0.57	8.18*	1.38
τ_3	17.62*	0.13	14.53*	0.45
τ_4	27.13*	0.70	22.71*	1.83
τ_5	40.08*	1.74	33.02*	3.26
τ_6	0.44	0.23	0.12	0.19
DW	2.12	1.74	2.01	1.74
μ_3	2.03	2.69*	2.19	4.08*
τ_7	6.09	5.78	5.26	5.83
μ_4			3.65	0.30
μ_5			0.33	0.45
μ_6	4.24*	0.40	1.69	0.28
τ_8	15.41*	3.25	12.73*	1.62

Notes: * significant at the 5% critical value.
For the meaning of the symbols used for the test statistics and the relative degrees of freedom, see table 1.

8. CONCLUDING REMARKS

At the end of our analysis, some concluding remarks may be very useful in providing some suggestions about the possibility of extending the research to cover more general cases than what we have done here.

The Monte Carlo simulation we have performed may already be considered as a slight generalization of GP, for it explicitly takes into account the presence of an intercept term in both competing models. The size and power estimates we have shown are not so different from GP's ones and this is particularly important for the adjusted Cox tests they propose. If we confine ourselves to the set up illustrated in section 5, further extensions are possible by considering different ways of generating the X variables (and consequently the Z variables) and the disturbance U. GP present several sets of experiments but they do not analyze what happens to the small sample behaviour of the tests in the dynamic case when the X variables and/or the vector of disturbances are not assumed to be normally distributed.

An area where interesting results may be obtained concerns the case where the two competing models are characterized by different (non nested) dynamics. This is a very important extension of the present work. In other terms, the problem is to find out what is the small sample behaviour of the usual non nested testing procedures when, for instance we have:

$$[8.1] \qquad M_0^* : y_t = \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

$$M_i^* : y_t = \alpha_0 x_t + \alpha_1 y_{t-1} + w_t$$

[8.1] is one of the several cases that may arise from the application of "general to specific" procedures. As it is well known there does not exist a unique ordering from the more general model to the more parsimonious representation we can find. In these cases, when two non nested final forms are regarded as plausible, then a possibility of discriminating between them may be given by

the application of non nested tests. Clearly, this is correct only if we have at least a rough idea of the small sample behaviour of the non nested testing procedures we want to apply.

Another possible issue we have raised in our analysis is the problem of the kind of size and power estimates we are actually computing (see section 6). This is a very important problem which needs further investigation. Basically, the issue is to establish a set of conditions by which the computation of the estimated size and power is symmetric under different DGP's. In so doing, a lot of computation and time may be saved on applying Monte Carlo simulations to non nested tests. In section 7.1 we have just offered some kind of evidence about a possible solution in particular cases (different number of regressors in the competing models). Clearly, more analytical and empirical work is required, especially when the competing models are specified as in [8.1].

Finally, the use of the response surface analysis has proved to be very efficient in simulation exercises such as our Monte Carlo study. Though it can be regarded only as a first step toward a more complete analysis - for instance by considering sequence of local alternatives - it allows to avoid the main limit of Monte Carlo simulations, that is their specificity. In our case, this means that the response surface we have obtained can be seen as an approximation to the small sample distribution of the test statistics under examination.

REFERENCES

Atkinson, A.C. (1970), "A method for discriminating between models", Journal of Royal Statistical Society, Series B, 32, 323-45.

Bera, A.K. and McAleer, M. (1985), "On exact and asymptotic tests of non nested models", Australian National University (ANU), Working Papers in Economics and Econometrics, n. 117.

Chow, G.C. (1960), "Tests of equality between sets of coefficients in two linear regressions", Econometrica, 28, 3, 591-605.

Cochran, W.G. and Cox, G.M. (1957), "Experimental designs", 2nd edition, New York, John Wiley & Sons.

Cox, D.R. (1961), "Tests of separate families of hypotheses", in J. Neyman (ed.) 'Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability', 1 Berkeley, University of California Press, 105-123.

——— (1962), "Further results on tests of separate families of hypotheses", Journal of the Royal Statistical Society, B,24, 406-424.

——— (1970), "The analysis of binary data", London,Methuen.

Das Gupta, S. and Perlman, M.D. (1974), "Power of the non central F test: effects of additional variates on Hotelling's t^2 test", Journal of American Statistical Association, 69, 174-180.

Dastoor, N.K. and McAleer, M. (1985), "Some comparisons of joint and paired tests of non nested hypotheses", ANU, Working Papers in Economics and Econometrics, n. 119.

Davidson, R. and McKinnon, J.G. (1981), "Several tests of model specification in the presence of alternative hypotheses", Econometrica, 49, 781-793.

——— and ——— (1982), "Some non nested hypotheses tests and the relations among them", Review of Economic Studies, 49, 551-565.

Durbin, J. and Watson, G.S. (1950), "Testing for serial correlation in least squares regressions. I", Biometrika, 37, 3 and 4, 409-428.

——— and ——— (1951), "Testing for serial correlation in least squares regressions. II", Biometrika, 38, 1 and 2, 159-178.

Engle, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", Econometrica, 50, 4, 987-1007.

Ericcson, N.R. (1983), "Asymptotic properties of instrumental variables statistics for testing non nested hypotheses", Review of Economic Studies, 50, 287-304.

——— (1986), "Post simulation analysis of Monte Carlo experiment: interpreting Pesaran's (1974) study of non nested hypotheses test statistics", Federal Reserve Bank, International Finance Discussion Papers, n. 276.

Fisher, G. and McAleer, M. (1979), "On the interpretation of the Cox test in econometrics", Economics Letters, 4, 145-150.

——— and ——— (1981), "Alternative procedures and associated tests of significance for non nested hypotheses", Journal of Econometrics, 16, 103-119.

Godfrey, L.G. and Pesaran, M.H. (1983), "Tests of non nested regression models: small sample adjustments and Monte Carlo evidence", Journal of Econometrics, 21, 133-154.

Harvey, A.C. (1981), "The econometric analysis of time series", Oxford, Philip Allan.

Hendry, D.F. (1979), "Predictive failure and econometric modelling in macroeconomics; the transaction demand for money", in P. Ormerod (ed.) 'Economic modelling', London, Heinemann Educational Books, 217-242.

——— (1984), "Monte Carlo experimentation in econometrics", in Z. Griliches and M.D. Intriligator (eds.) 'Handbook of econometrics', vol. II, ch. 16, Amsterdam, North-Holland, 937-976.

Johnston, J. (1963), "Econometric methods", New York, McGraw-Hill.

Kendall, M.G. (1946), "The advanced theory of statistics", vol. II, London, Charles Griffin.

Kiefer, N.M. and Salmon, M. (1983), "Testing normality in econometric models", Economics Letters, 11.

McAleer, M. (1984), "Specification tests for separate models", ANU, Working Papers in Economics and Econometrics, n. 110.

——— and Pesaran, M.H. (1985), "Non nested tests: a survey", ANU, Working Papers in Economics and Econometrics, n. 128.

McKinnon, J.G. (1983), "Model specification tests against non nested alternatives", Econometric Review, 2, 1, 85-110.

Mizon, G.E. (1984), "The encompassing approach in econometrics", in D.F. Hendry and K.F. Wallis (eds.) 'Econometrics and quantitative economics', ch. 6, 135-172, Oxford, Blackwell.

——— and Hendry, D.F. (1980), "an empirical application and Monte Carlo analysis of dynamic specification", Review of Economic Studies, 47, 1, 21-45.

——— and Richard, J.F. (1986), "The encompassing principle and its applications to testing non nested hypotheses", Econometrica, 54, 3, 657-678.

Pesaran, M.H. (1974), "On the general problem of model selection", Review of Economic Studies, 41, 2, 153-171.

——— (1982a), "Comparisons of local power alternative tests of non nested regression models", Econometrica, 50, 1287-1305.

——— (1982b), "On the comprehensive method of testing non nested regression models", Journal of Econometrics, 18, 263-274.

White, H. (1980), Using least squares to approximate unknown regression functions", International Economic Review, 21, 1, 149-170.

RECENTLY PUBLISHED «TEMI» (*)

n. 91 — *La disoccupazione in Italia: un'analisi con il modello econometrico della Banca d'Italia*, di G. Bodo - I. Visco (luglio 1987).

n. 92 — *L'Italia e il sistema monetario internazionale dagli anni '60 agli anni '90 del secolo scorso*, di M. Roccas (agosto 1987).

n. 93 — *Reddito e disoccupazione negli Stati Uniti e in Europa: 1979-1985*, di J. C. Martinez Oliva (agosto 1987).

n. 94 — *La tassazione e i mercati finanziari*, di G. Ancidoni - B. Bianchi - V. Ceriani - P. Coraggio - A. Di Majo - R. Marcelli - N. Pietrafesa (agosto 1987).

n. 95 — *Una applicazione del filtro di Kalman per la previsione dei depositi bancari*, di A. Cividini - C. Cottarelli (ottobre 1987).

n. 96 — *Macroeconomic Policy Coordination of Interdependent Economies: the Game-Theory Approach in a Static Framework*, by J. C. Martinez Oliva (ottobre 1987).

n. 97 — *Occupazione e disoccupazione: tendenze di fondo e variazioni di breve periodo*, di P. Sylos Labini (novembre 1987).

n. 98 — *Capital controls and bank regulation*, by G. Gennotte - D. Pyle (dicembre 1987).

n. 99 — *Funzioni di costo e obiettivi di efficienza nella produzione bancaria*, di G. Lanciotti - T. Raganelli (febbraio 1988).

n. 100 — *L'imputazione di informazioni mancanti: una sperimentazione*, di L. Cannari (marzo 1988).

n. 101 — *Esiste una curva di Beveridge per l'Italia? Analisi empiriche della relazione tra disoccupazione e posti di lavoro vacanti (1980-1985)*, di P. Sestito (marzo 1988).

n. 102 — *Alcune considerazioni sugli effetti di capitalizzazione determinati dalla tassazione dei titoli di Stato*, di D. Franco - N. Sartor (luglio 1988).

n. 103 — *La coesione dello SME e il ruolo dei fattori esterni: un'analisi in termini di commercio estero*, di L. Bini Smaghi - S. Vona (luglio 1988).

n. 104 — *Stime in tempo reale della produzione industriale*, di G. Bodo - A. Cividini - L. F. Signorini (luglio 1988).

n. 105 — *On the difference between tax and spending policies in models with finite horizons*, by W. H. Branson - G. Galli (ottobre 1988).

(*) Requests for copies should be sent to: Banca d'Italia - Servizio Studi, Biblioteca, via Nazionale 91, 00184 Rome.

BANCA D'ITALIA - CENTRO STAMPA